8.3-2

Test Video Selection for Video Processing Research and Development

Chin Chye Koh[1], Jorge E. Caviedes[2], and Sanjit K.Mitra[3].
[1] Video Products Group, Inc., Oxnard, CA, USA; [2] Intel Corporation, USA; [3] University of California Santa Barbara, USA.

***Abstract*--Many video processing algorithms have been developed in recent years. However, visual performance validation of these algorithms is often overlooked. We discuss the need for a standardized suite of free test videos appropriately selected according to a set of usage rules. These videos are to be stored in a publicly accessible online library that will facilitate effective collaboration and interaction among research scientists and development engineers.**

I. INTRODUCTION

In the consumer electronics R&D environment, the development and integration of metrics and algorithms is defined by a relatively rigid work flow. At each stage of the development, validation is required to ensure that the system meets the required visual quality. These tests require the selection of test video content under strict guidelines. Test video selection is a complex problem and is generally hampered by: (1) lack of expertise in selection of test videos, (2) videos selected may be best choice available and not relevant to the video space targeted, and (3) lack of availability of relevant videos. These problems create difficulties in the proper benchmarking and comparison of new technologies and products.

Video content selection and the associated criteria can be illustrated pictorially in the figure below. The linked circles represent the three fundamental components in video R&D that are to be accounted for.

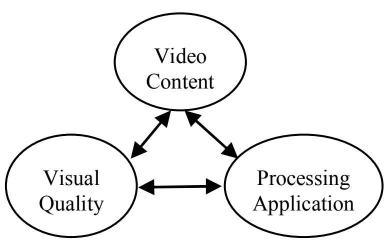

Fig. 1. Interrelation between the three components of video content, process type, and visual quality

Video Content represents the need to use a complete and representative set of test videos for the purpose of visual quality validation. Visual quality must be managed from the early stages of the development cycle, including early exploratory research.

Processing Application represents the need to consider the type of video processing under test. For example, content for evaluating compression may not be appropriate for sharpness enhancement.

Visual Quality represents the need to evaluate the specific algorithm performance with tests that check for perceived visual quality. These tests may be subjective (psychophysical experiments) or objective metrics that correlate well with the human visual system (HVS) or the mean subjective scores from a standardized test.

Our solution for the selection of test video content is framed within the Consumer Digital Video Library (CDVL) [1, 2]. The CDVL is a collaborative project between several interested parties in the development of a standardized set of video content for the specific purpose of supporting video processing R&D. The goal of the CDVL is to provide a repository of test video content and a set of usage rules. The video content would be housed in a central location and interfaced via an online virtual digital library.

In the following sections we discuss the usage rules that guide test content selection, the content descriptors that describe the test videos, the prototype CDVL, and close with some concluding remarks.

II. USAGE RULES

The purpose of defining usage rules is to create a framework in which users can select appropriate test video content suitable for the ultimate visual quality objectives. These usage rules are parameterized by several factors including process type, developmental stage, and validation type.

A. Processes

The processes that are of interest can be categorized into three broad areas that account for most of the types of processes found in consumer products: corrective (attenuation of defects), enhancement (increase visual quality), and conditioning (conversion between formats). There is no single video or test pattern that can be used effectively across these categories. Thus, test video content must be selected according to the process type. In video processing R&D, there are two important points to take note of, the ***content requirements*** and the ***performance criteria***.

For each process type, the ***content requirements*** can be very different. Thus, it is necessary to define the types of video content required for a specific process. Corrective algorithms require video contents that are of medium to high spatial, temporal, and color complexities, including outdoor scenes with textures/details, and close-ups of people. These videos should contain the artifact to a noticeable level but to the extent that it is repairable. Enhancement algorithms require

This work was supported by a University of California MICRO grant with matching support from Intel Corporation.

video content that contain high spatial and color complexity combined with low to medium temporal complexity. Additionally, the content should be mostly unimpaired and clean, but responsive to different types of enhancements that typically would make good content look better. Conditioning processes require video content that contain high spatial and color complexity combined with cases including low to high temporal complexity. Additionally, the content should be mostly unimpaired and clean. Finally, all processes are tested with a mix of video content that contains the levels of impairments representative of the consumer video space.

The ***performance criterion*** for each process type can also be somewhat different. However all process require visual confirmation of successful correction, enhancement or conditioning without introducing new annoyance factors. For correction and enhancement, the processed image must be preferred at a specified level in a subjective test. Expert tests must also corroborate effectiveness of algorithm. For conditioning, visual confirmation of the processing should look for the retention of proper motion portrayal and color integrity and that small objects in the video content should not be removed by the processing.

B. Developmental Stage

The various stages in the research and development of a product or algorithm can be classified into the following stages, exploration (conceptualization), prototyping (proof of concept), integration (to ensure that algorithm works as part of a larger system), and production (system verification before deployment). Each of these stages may require different types and numbers of videos. For example, during the exploratory stage, only a small number of critical videos are necessary as the time frame is short and the validation requirements are less stringent. In the integration phase, a large number of videos of longer duration are required to benchmark the system to ensure that all functions operate as expected.

C. Test Type

The validation test can be of the subjective or the objective type. Subjective test involves psychophysical experiments that evaluate the effectiveness of an algorithm with human subjects. Objective tests are metric based tests where processed videos are measured using a metric and benchmarked. These objective metrics may consist of traditional metrics such as PSNR [3, 4] or metrics designed to correlate well with the human visual system (HVS) [3, 4].

III. CONTENT DESCRIPTORS

Content descriptors allow the categorization of videos by physical attributes. These attributes can then be used in the selection of video content with the usage rules as a guide. We have defined three content descriptors: color richness, spatial complexity, and temporal complexity, each divided into low, medium, and high levels. Notice that these descriptors are still in a developmental stage and the metrics proposed below will require further enhancement.

Color richness was computed as the product of the standard deviations along the three perceptual axes, L^*, u^*, and v^* in the approximately perceptually uniform CIE 1976 ($L^*u^*v^*$) color–space [5]. Spatial complexity was defined as a combination of entropy [6], a blurriness metric [7], and a spatial activity metric [8]. Temporal complexity was computed as the product of the mean, median, standard deviation, and maximum value after removing the top 5% of the motion vector magnitudes [9]. All metrics were computed for individual frames and summed together to form the frame metric and finally averaged across all frames to yield a single video metric.

IV. CONCLUDING REMARKS

In this paper, we have discussed the need for a standardized test suite of video content and the definition of a set of usage rules for the selection of video content. Our solution is the implementation of a Consumer Digital Video Library (CDVL) that consists of a set of usage rules in which specific test videos are selected depending on the application under study. The CDVL in prototype form is presently undergoing alpha test. The CDVL is expected to go into public beta test in the fourth quarter of 2007. Our near term goals are to develop the content descriptors described in this work and to populate the CDVL with a number of highly relevant videos and the related usage rules. Our long term goal is to push on towards standardization of test videos and the associated methodology.

EXAMPLES OF REFERENCE STYLES

[1] C.C. Koh and S.K. Mitra, “A proposed framework for a digital video library for research applications,” *Proc. of the 2nd VPQM, Scottsdale*, AZ, Jan 2006.

[2] C.C. Koh, J.E. Caviedes, and S.K. Mitra, “Content management in a consumer digital video library design,” *Proc. of the 3rd VPQM*, Scottsdale, AZ, Jan 2007.

[3] Z. Wang, H.R. Sheikh, and A.C. Bovik, “Objective video quality assessment,” *The Handbook of Video Databases: Design and Applications*, B. Furht and O. Marqure, Ed. CRC Press, pp. 1041–1078, Boca Raton, FL, Sep 2003.

[4] S. Winkler, “Video quality metrics – A review,” *Digital Video Image Quality & Perceptual Coding*, H.R. Wu and K.R. Rao, Eds. CRC Press, pp. 155–180, Nov 2005.

[5] G. Wyszecki and W.S. Styles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., John Wiley & Sons, Inc., New York, 1982.

[6] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, 2nd Edition, Prentice Hall, Upper Saddle River, NJ, 2002.

[7] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, “Perceptual blur and ringing metrics: Application to JPEG2000,” *Signal Processing: Image Communication*, vol. 19, iss. 2 , pp. 163–172, Feb 2004,

[8] S. Wolf and M.H. Pinson, “Spatial–temporal distortion metrics for in-service quality monitoring of any digital video system,” *SPIE Proc., Multimedia Systems and Applications II*, vol. 3845, Sep. 1999, pp. 266–277.

[9] K.A. Peker, A. Divakaran, and T.V. Papathomas, “Automatic measurement of intensity of motion activity of video segments,” *SPIE Proc., Storage and Retrieval for Media Databases*, vol. 4315, San Jose, CA, 2001, pp. 341-351.